THE COMBINATION OF TIME SERIES AND CROSS-SECTION DATA IN INTERINDUSTRY FLOW ANALYSIS

Kenneth J. Arrow

Stanford University

P-942

18 September 1956

The RAND Corporation

1700 MAIN ST. SANTA MONICA, CALIFORNIA

CLEARINGHOUSE FOR FEDERAL SCIENTIFIC AND TECHNICAL INFORMATION CFSTI
DOCUMENT MANAGEMENT BRANCH 410 11

LIMITATIONS IN REPRODUCTION QUALITY

ACCESSION ::

☒ 1 WE REGRET THAT LEGIBILITY OF THIS DOCUMENT IS IN PART UNSATISFACTORY REPRODUCTION HAS BEEN MADE FROM BEST AVAILABLE COPY

☐ 2 A PORTION OF THE ORIGINAL DOCUMENT CONTAINS FINE DETAIL WHICH MAY MAKE READING OF PHOTOCOPY DIFFICULT

☐ 3 THE ORIGINAL DOCUMENT CONTAINS COLOR, BUT DISTRIBUTION COPIES ARE AVAILABLE IN BLACK-AND-WHITE REPRODUCTION ONLY

☐ 4 THE INITIAL DISTRIBUTION COPIES CONTAIN COLOR WHICH WILL BE SHOWN IN BLACK-AND-WHITE WHEN IT IS NECESSARY TO REPRINT

☐ 5 LIMITED SUPPLY ON HAND WHEN EXHAUSTED DOCUMENT WILL BE AVAILABLE IN MICROFICHE ONLY

☐ 6 LIMITED SUPPLY ON HAND WHEN EXHAUSTED DOCUMENT WILL NOT BE AVAILABLE

☐ 7 DOCUMENT IS AVAILABLE IN MICROFICHE ONLY

☐ 8 DOCUMENT AVAILABLE ON LOAN FROM CFSTI (TT DOCUMENTS ONLY)

☐ 9

PROCESSOR

TSL-107-10 63

THE COMBINATION OF TIME SERIES AND CROSS-SECTION DATA

IN INTERINDUSTRY FLOW ANALYSIS*

Kenneth J. Arrow
Stanford University

0. Summary

The following problem arose in the course of a larger study which sought to explain the variations of input-output ratios over time. (By the input-output ratio of industry i to industry j is meant the ratio of that part of the output of industry i used by industry j to the output of industry j; (see [1]). For such a study, there are two types of data available. For all years, there are available (ideally) outputs and final demands (the final demand for an industry consists of all uses of its product other than in other industries or itself) for all industries. The "balance equations" of input-output analysis (see [3] below) form an (incomplete) system of simultaneous relations which may be estimated by some version of the method of maximum likelihood (for computational reasons, the single-equation limited-information method is the only one likely to be used). However, for some years, we have additional information in the form of knowing the actual interindustry flows, which clearly should substantially increase the accuracy of our estimates.

The simplest technique is, of course, to assume that the "true" input-output ratio for any year for which flow data are available is exactly equal to the observed input-output ratio for that year. The assumption behind this is, however, contradictory to the basic postulate that all the relations involved are valid only up to a stochastic term.

It is therefore of interest to consider more explicitly the interindustry flow model implied in the use of both time series and interindustry flow data for estimating the parameters involved. The maximum likelihood estimates for a single

* This paper arose out of a study conducted by the RAND Corporation. I am indebted to Ronald W. Shephard, now of the Sandia Corporation, for raising the problem.

equation are then derived, in the sense of being expressed as the solution of a system of simultaneous equations. Unfortunately, these equations are rather cumbersome, and it is hoped that some reasonably efficient method of soultion can be found.

1. A Model With Constant Input-Output Coefficients

For expository reasons, we will start by assuming that the input-output ratios are constant over time. We consider the estimation of a single one of the balance equations of the input-output table. Let

x_{ot} = derived demand for commodity 0 at time *t*, i.e., total net output less final demand,

x_{jt} = output of commodity *j* at time *t* for $j = 1, \ldots, N$,

$\bar{x}_{jt}$ = input of commodity 0 to industry *j* at time *t*.

We have the identity,

$$(1) \quad x_{ot} = \sum_{j=1}^{N} \bar{x}_{jt},$$

where N is the number of commodities other than commodity 0. The assumption of input-output coefficients constant over time can be written,

$$(2) \quad \bar{x}_{jt} = \alpha_j x_{jt} + u_{jt},$$

where u_{jt} is a disturbance, distributed normally with mean zero. From (1) and (2), we have the balance equation for industry 0,

$$(3) \quad x_{ot} = \sum_{j=1}^{N} \alpha_j x_{jt} + u_t,$$

where

$$(4) \quad u = \sum_{j=1}^{N} u_{jt}.$$

The disturbance u_t is therefore normally distributed with mean 0.

It is assumed that there are two types of observations, one in which only the variables x_{jt} $(j = 0, \ldots, N)$ are observed and one in which the variables $\bar{x}_{jt}$ are also observed. Let U be the set of years of the first type, V the set of years of the second type; in application, V contains only the year 1947 plus possibly the year 1939. Let

$\bar{x}_{N+j,t} = x_{jt}$ for any $t \in V$,

x_t be the column vector with components x_{jt} $(j = 0, \ldots, N)$,

$\bar{x}_t$ the column vector with components $\bar{x}_{jt}$ $(j = 1, \ldots, 2N)$,

z_t the column vector of the predetermined variables,

π the regression matrix of x_t on z_t,

$\bar{\pi}$ the regression matrix of $\bar{x}_t$ on z_t.

The reduced forms for x_t, $\bar{x}_t$ can therefore be written,

(5) $x_t = \pi z_t + v_t$,

(6) $\bar{x}_t = \bar{\pi} z_t + \bar{v}_t$,

where the vectors v_t, $\bar{v}_t$ are each distributed normally with means zero and independently of z_t.

Let π_j be the j^{th} row of π, $\bar{\pi}_j$ the j^{th} row of $\bar{\pi}$. By substituting (5) and (6) into (1) and (2) and equating coefficients of z_t in the usual way (see [2]),

(7) $\pi_0 = \sum_{j=1}^{N} \bar{\pi}_j$,

(8) $\bar{\pi}_j = \alpha_j \pi_j$ $(j = 1, \ldots, N)$.

From the definition of $\bar{x}_{N+j,t}$,

(9) $\pi_j = \bar{\pi}_{N+j}$.

Let Λ be the covariance matrix of v_t, $\bar{\Lambda}$ that of $\bar{v}_t$. Then the distributions

of x_t and $\bar{x}_t$ (for given z_t) are given by

$$(10)\quad f(x_t) = k_1 |\Lambda|^{\frac{1}{2}} \exp\left[-(\tfrac{1}{2})(x_t - \pi z_t)' \Lambda (x_t - \pi z_t)\right],$$

$$(11)\quad g(\bar{x}_t) = k_2 |\bar{\Lambda}|^{\frac{1}{2}} \exp\left[-(\tfrac{1}{2})(\bar{x}_t - \bar{\pi} z_t)' \bar{\Lambda} (\bar{x}_t - \bar{\pi} z_t)\right],$$

where k_1 and k_2 are constants. For t in U, x_t is observed; for t in V, $\bar{x}_t$ is observed. Hence, the likelihood function L is obtained (assuming serial independence of the disturbances) by multiplying together all expressions of form (10) with $t \in U$ with all expressions of form (11) with $t \in V$.

$$(12)\quad \log L = C + (T/2) \log |\Lambda| - (\tfrac{1}{2}) \sum_{t \in U} (x_t - \pi z_t)' \Lambda (x_t - \pi z_t)$$
$$+ (S/2) \log |\bar{\Lambda}| - (\tfrac{1}{2}) \sum_{s \in V} (\bar{x}_s - \bar{\pi} z_t)' \bar{\Lambda} (\bar{x}_s - \bar{\pi} z_t),$$

where T and S are numbers of observations in U and V, respectively.

The aim, then, is to maximize log L with respect to π, $\bar{\pi}$, $\alpha_1, \ldots, \alpha_N$, Λ, and $\bar{\Lambda}$, subject to the restraints (7-9). The resulting values of $\alpha_1, \ldots, \alpha_N$ are the required maximum likelihood estimates.

2. The Model With Varying Input-Output Ratios

In general, each input-output ratio is presumed to vary linearly with a number of other variables. Let $w_1, \ldots, w_K$ be the variables on which the input-output ratios α_j depend. Any one coefficient may depend on only some of these variables; let K_j be the set of variables on which α_j depends. For simplicity of notation, let w_0 be the constant 1, and assume that K_j contains

0 for all j.

(13) $\alpha_j = \sum_{k \in K_j} \alpha_{kj} w_k$.

(13) and (2) can be written,

(14) $\bar{x}_{jt} = (\sum_{k \in K_j} \alpha_{kj} w_{kt}) x_{jt} + u_{jt}$ $(j = 1, \ldots, N)$.

Let

$$x_{Nk+j,t} = w_{kt} x_{jt} \quad (k = 0, \ldots, K;\ j = 1, \ldots, N).$$

For $k = 0$, this definition corresponds to the earlier one, since $w_{ot} = 1$ for all t. Also, let

(15) $\bar{x}_{Nk+j,t} = x_{N(k-1)+j,t}$ $(k = 1, \ldots, K+1;\ j = 1, \ldots, N)$,

x_t be the column vector whose components are x_{jt} $(j = 0, \ldots, N(K+1))$,

$\bar{x}_t$ be the column vector whose components are $\bar{x}_{jt}$ $(j = 1, \ldots, N(K+2))$.

The vectors x_t and $\bar{x}_t$ are the observables in the observations under U and V, respectively; they are somewhat redefined from section 1. (14) can be written,

(16) $\bar{x}_{jt} = \sum_{k \in K_j} \alpha_{kj} x_{Nk+j,t}$ $(j = 1, \ldots, N)$.

Define π and $\bar{\pi}$, as before, as the regression matrices of x_t and $\bar{x}_t$, respectively, on the vector of predetermined variables z_t. Since most of the components of the two vectors are formed as a product of two variables, there is no reason for the regressions to be linear; however, as Anderson and Rubin have shown, acting as if they were linear normal leads to consistent estimates (see [3], p. 574). Hence, equations (5) and (6) will still be regarded as valid. Since (1) still holds, (7) is still valid. Substituting (5) and

(6) into (16) yields the following analogue of (8):

$$(17)\quad \bar{\pi}_j = \sum_{k \in K_j} \alpha_{kj} \pi_{Nk+j} \qquad (j = 1, \ldots, N).$$

Finally, (15) implies that,

$$(18)\quad \bar{\pi}_{Nk+j} = \pi_{N(k-1)+j} \quad (k = 1, \ldots, K+1;\ j = 1, \ldots, N).$$

The likelihood function has the same form as before. Hence the maximum likelihood estimates are obtained by maximizing (12) subject to the restraints (7), (17), and (18).

3. Restrictions on the Covariance Matrices of the Reduced Form Disturbances.

It can be argued that the above model implies certain relations between the covariance matrices of the reduced forms (5) and (6), i.e., Λ^{-1} and $\bar{\Lambda}^{-1}$. Since (15) is an identity,

$$(19)\quad \bar{v}_{Nk+j,t} = v_{N(k-1)+j,t} \quad (k = 1, \ldots, K+1;\ j = 1, \ldots, N)$$

Let σ_{ij} be the covariance of v_i and v_j, $\bar{\sigma}_{ij}$ the covariance of $\bar{v}_i$ and $\bar{v}_j$. Then, from (19)

$$\bar{\sigma}_{Nk+j,\ Nk'+j'} = \sigma_{N(k-1)+j,\ N(k'-1)+j'} \quad (k,k' = 1, \ldots, K+1;\ j,j' = 1, \ldots, N)$$

which can also be written

$$(20)\quad \bar{\sigma}_{ij} = \sigma_{i-N,j-N} \quad (i = N+1, \ldots, N(K+2);\ j = N+1, \ldots, N(K+2)).$$

Similarly, since (1) is an identity,

$$(21)\quad \sigma_{oi} = \sum_{j=1}^{N} \bar{\sigma}_{j,i+N} \quad (i = 1, \ldots, N(K+1)).$$

It would be reasonable to impose the conditions (20) and (21) upon the maximization of the likelihood function. However, this would greatly complicate

the form of the estimation procedure. Further, if we wish to assume that the interindustry flows are observed only with error, the relation (1) becomes a stochastic relation (if $\bar{x}_{jt}$ is taken to refer to the observed rather than the actual flow), and the argument leading to (21) breaks down. For these reasons, we will disregard the restrictions on the covariance matrices of the reduced form disturbances.

4. The Derivation of the Estimates

Rewrite (18) as

(22) $\bar{\pi}_j = \pi_{j-N} \quad (j = N+1, \ldots, N(K+2)).$

We wish to maximize (12) subject to (7), (17), and (22). We use the method of Lagrange multipliers. There are altogether $N(K+2)+1$ restrictions, each a vector restriction with as many components as there are predetermined variables. To each restriction we assign a Lagrange multiplier, itself a vector. Let λ correspond to (7), μ^j to the j^{th} restriction in (17), and ν^j to the j^{th} restriction in (22). The Lagrangian may then be written,

$$(23) \quad A = C + (T/2) \log |\Lambda| + (S/2) \log |\bar{\Lambda}| - (1/2) \sum_{t \in U} (x_t - \pi z_t)' \Lambda (x_t - \pi z_t)$$

$$- (1/2) \sum_{s \in V} (\bar{x}_s - \bar{\pi} z_s)' \bar{\Lambda} (\bar{x}_s - \bar{\pi} z_s)$$

$$+ (\pi_o - \sum_{j=1}^{N} \bar{\pi}_j) \lambda + \sum_{j=1}^{N} (\sum_{k \in K_j} \alpha_{kj} \pi_{Nk+j} - \bar{\pi}_j) \mu^j$$

$$+ \sum_{j=N+1}^{N(K+2)} (\pi_{j-N} - \bar{\pi}_j) \nu^j.$$

Let M_{xz} be the matrix of sums of cross-products of x and z, summation extending over all observations in U,

M_{zz} be the matrix of sums of squares and cross-products of the z's, summed over observations in U,

$\bar{M}_{xz}$ be the matrix of sums of cross-products of $\bar{x}$ and z over V,

$\bar{M}_{zz}$ be the matrix of sums of squares and cross-products of the z's, summed over V.

Now differentiate A successively with respect to π_o, π_{Nk+j} ($k = 0, \ldots, K$; $j = 1, \ldots, N$), $\bar{\pi}_j$, and equate the results to 0.

(24) $\partial A/\pi_o = \Lambda^0 (M_{xz} - \pi M_{zz}) + \lambda' = 0.$

(25) $\partial A/\partial \pi_{Nk+j} = \Lambda^{Nk+j} (M_{xz} - \pi M_{zz}) + \alpha_{kj} \mu^{j'} + \nu^{N(k+1)+j'} = 0,$

$(k = 0, \ldots, K;\ j = 1, \ldots, N).$

(26) $\partial A/\bar{\pi}_j = \bar{\Lambda}^j (\bar{M}_{xz} - \bar{\pi} \bar{M}_{zz}) - \mu^{j'} - \lambda' = 0 \quad (j = 1, \ldots, N).$

(27) $\partial A/\bar{\pi}_j = \bar{\Lambda}^j (\bar{M}_{xz} - \bar{\pi} \bar{M}_{zz}) - \nu^{j'} = 0 \quad (j = N+1, \ldots, N(K+2)).$

Here, each equation represents a row vector of derivatives. In (25), it is understood that

(28) $\alpha_{kj} = 0$ if $k \notin K_j$.

Define row vectors,

(29) $P_o = -\lambda'$, $P_{Nk+j} = -(\alpha_{kj} \mu^{j'} + \nu^{N(k+1)+j'})$ $(k = 0, \ldots, K;\ j = 1, \ldots, N)$,

(30) $R_j = \lambda' + \mu^{j'}$ $(j = 1, \ldots, N)$, $R_j = \nu^{j'}$ $(j = N+1, \ldots, N(K+2))$,

and matrices,

(31) $$P = \begin{bmatrix} P_o \\ P_1 \\ \vdots \\ P_{N(K+1)} \end{bmatrix}, \quad R = \begin{bmatrix} R_o \\ R_1 \\ \vdots \\ R_{N(K+2)} \end{bmatrix}.$$

Then (24) and (25) may be written,

(32) $\Lambda(M_{xz} - \pi M_{zz}) = P,$

(26) and (27) become

(33) $\bar{\Lambda}(\bar{M}_{\bar{x}z} - \bar{\pi}\bar{M}_{zz}) = R.$

From (29) and (30),

(34) $P_{Nk+j} = -\alpha_{kj} R_j - R_{N(k+1)+j} - \alpha_{kj} P_o .$

Define

(35) $Q_{oq} = 0,\ Q_{Nk+j,j} = -\alpha_{kj},\ Q_{Nk+j,N(k+1)+j} = -1,\ Q_{Nk+j,q} = 0$ for all $q \neq j$,

$N(k+1)+j\ (q = 1, \ldots, N(K+2);\ k = 0, \ldots, K;\ j = 1, \ldots, N).$

Let Q be the matrix with elements Q_{pq} $(p = 0, \ldots, N(K+1);\ q = 1, \ldots, N(K+2))$.
Let $\bar{Q}$ be the column vector with $\bar{Q}_o = 1,\ \bar{Q}_{Nk+j} = -\alpha_{kj}$ $(k = 0, \ldots, K;\ j = 1, \ldots, N)$.

Then (34) can be written,

(36) $P = Q\,R + \bar{Q}\,P_o .$

If we combine (32) and (36),

(37) $\Lambda(M_{xz} - \underline{\pi} M_{zz}) = Q\,R + \bar{Q}\,P_o .$

Note that Q and $\bar{Q}$ depend only on the α_{kj}'s. In effect, the unknown Lagrange parameters λ, μ^j $(j = 1, \ldots, N)$, ν^j $(j = N+1, \ldots, N(K+2))$ have been transformed into the unknown matrix R and vector P_o. (33) and (37) are the results of the differentiation of the Lagrangian (23) with respect to the regression matrices π, $\bar{\pi}$. Now differentiate with respect to α_{kj}.

(38) $\pi_{Nk+j}\,\mu^j = 0$ $(k = 0, \ldots, K;\ j = 1, \ldots, N)$.

With the aid of (29) and (30), (38) may be rewritten

(39) $\pi_{Nk+j}\,(R_j + P_o)' = 0$ $(k = 0, \ldots, K;\ j = 1, \ldots, N)$.

We can eliminate P_o as follows: In (37), multiply on the left by $\bar{Q}' \Sigma$, where $\Sigma = \Lambda^{-1}$. Let

(40) $\gamma = (\bar{Q}' \Sigma \bar{Q})^{-1}$.

Note that γ is a scalar. Then,

(41) $P_o = \gamma (\bar{Q}' M_{xz} - \bar{Q}' \pi M_{zz} - \bar{Q}' \Sigma \; Q R)$.

Finally, we may differentiate A with respect to the matrices Λ, $\bar{\Lambda}$, the inverses of the covariance matrices of reduced form disturbances in the equations in the observations of U and V, respectively. Let $\bar{\Sigma} = \bar{\Lambda}^{-1}$. Then, as in Anderson and Rubin (see n. 2),

(42) $\Sigma = M_{xx} - M_{xz} \; \pi' - \pi \; M_{zx} + \pi M_{zz} \; \pi'$;

(43) $\bar{\Sigma} = \bar{M}_{\bar{x}\bar{x}} - \bar{M}_{\bar{x}z} \; \bar{\pi}' - \bar{\pi} \; \bar{M}_{z\bar{x}} + \bar{\pi} \; \bar{M}_{zz} \; \bar{\pi}'$.

We have then to solve equations (33), (37), (39), (42), and (43) for the unknowns Λ, $\bar{\Lambda}$, $\bar{Q}$, R, γ, π, and $\bar{\pi}$, where P_o is eliminated with the aid of (40) and (41). Actually, we are only interested in $\bar{Q}$ (note that the matrix Q is completely defined by the vector $\bar{Q}$) which involves the structural coefficients to be estimated. These equations seem considerably more difficult to handle than the corresponding ones of Anderson and Rubin.

REFERENCES

[1] W. W. LEONTIEF, *The Structure of the American Economy: 1919-1939* (Second Edition) New York: Oxford University Press, 1951.

[2] T. W. ANDERSON AND H. RUBIN, "Estimation of the Parameters of a Single Equation in a Complete System of Stochastic Equations", *Annals of Mathematical Statistics*, Vol. 20 (1949), pp. 46-63.

[3] T. W. ANDERSON AND H. RUBIN, "The Asymptotic Properties of Estimates of the Parameters of a Single Equation in a Complete System of Stochastic Equations", *Annals of Mathematical Statistics*, Vol. 21 (1950), pp. 570-582.

KJA:gbd